April, 1956 BU-65-M

RANDOM MATING AMONG FULL SIBS, AN EXPERIMENTAL TECHNIQUE FOR ESTIMATING GENETIC VARIANCE COMPONENTS IN A HAPLOID POPULATION

Douglas S. Robson
Cornell University

Introduction

We shall show that random mating among full sibs in successive generations obtained from the cross of two haploid parents is an effective experimental procedure for estimating genetic variance components. Restricted random mating is, in fact, the basic procedure for quantitative inheritance studies involving two pure lines and must be followed, in one form or another, if the experiment is to yield estimates of all possible epistatic components of variance.

This paper is the result of a contemplation of the problem of designing an experiment for studying quantitative inheritance in the haploid organism Neurospora. We define all possible epistatic effects and, under the assumption of no linkage, compute the various genetic variances and covariances which arise under the successive random mating of full sibs. Estimation of these genetic variances and covariances then permits, in addition, estimation of all possible epistatic components of variance.

The Single Locus Problem

We begin by considering the simplest case, that of a cross between two haploids which differ genetically at a single locus. Let A_0 and A_1 denote the two alleles involved so that we may express the genotype of one parent as A_0 and the genotype of the other as A_1. We shall employ the same symbols, A_0 and A_1, to denote genotypic value (average phenotype with respect to the quantitative character) of the genotypes A_0 and A_1, respectively. The genotypic effect, a, will then be defined as the difference between the genotypic value A_0 and the average genotypic value $\frac{A_0 + A_1}{2}$; thus,

$$a = A_0 - \frac{A_0 + A_1}{2} = \frac{A_0 + A_1}{2} - A_1 = \frac{A_0 - A_1}{2}$$

or

$$A_0 = \frac{A_0 + A_1}{2} + a$$

$$A_1 = \frac{A_0 + A_1}{2} - a$$

or, letting

$$a_0 = \frac{A_0 + A_1}{2}$$

we have the identity

$$(1) \qquad A_j = a_0 + (1-2_j)a.$$

In the F_1 population obtained by crossing the genotypes A_0 and A_1 the genotypic distribution is

$$P_{F_1}\left\{ A_j = A_0 \right\} = \frac{1}{2} = P_{F_1}\left\{ A_j = A_1 \right\}.$$

The mean genotypic value in this population is therefore

$$E\left\{A_j\right\} = a_0 + aE\left\{1-2_j\right\} = a_0$$

and the genetic variance, denoted by V, is

$$V = a^2E\left\{1-2_j\right\}^2 = a^2.$$

Random mating among F_1 individuals yields three possible crosses, $A_0 \times A_0$ with probability $\frac{1}{4}$, $A_0 \times A_1$ with probability $\frac{1}{2}$, and $A_1 \times A_1$ with probability $\frac{1}{4}$. Each of these three crosses generates an "F_2" family. Thus, in the F_2 generation the family consisting entirely of the genotype A_j appears with probability $\frac{1}{4}$, $j = 0,1$, and the family consisting of $\frac{1}{2}$ genotype A_0 and $\frac{1}{2}$ genotype A_1 appears with probability $\frac{1}{2}$. The genotype A_j thus appears with frequency $\frac{1}{2}$, $j = 0,1$, in the F_2 population and in general, of course, random mating among members of a family will here simply reproduce the genotypic distribution in the family. The mean genotypic value among the progeny obtained by random mating within a family will therefore be equal to the parental mean in that

family. The F_2 mean is therefore equal to the F_1 mean a_0. The genetic variance among F_2 family means, denoted by V(2,1) is therefore

$$(2) \quad V(2,1) = \tfrac{1}{4}(a_0+a)^2 + \tfrac{1}{2}[\tfrac{1}{2}(a_0+a) + \tfrac{1}{2}(a_0-a)]^2 + \tfrac{1}{4}(a_0-a)^2 - a_0^2 = \tfrac{1}{2}a^2$$

and the average genetic variance within F_2 families, denoted by V(2,2) is then

$$V(2,2) = \tfrac{1}{4}(0) + \tfrac{1}{2}a^2 + \tfrac{1}{4}(0) = \tfrac{1}{2}a^2.$$

The total variance V in the F_2 population is the same as that in the F_1,

$$V = V(2,1) + V(2,2) = \tfrac{1}{2}a^2 + \tfrac{1}{2}a^2 = a^2.$$

Random mating among F_2 full sibs will again produce three genetically distinct F_3 families. Thus, random mating among members of the F_2 family $\{A_0\}$, which appears with frequency $\frac{1}{4}$ in the F_2 population, will yield the F_3 family $\{A_0\}$; similarly, random mating among members of the F_2 family $\{A_1\}$ will yield the F_3 family $\{A_1\}$; random mating among members of the F_2 family consisting of $\frac{1}{2}$ genotype A_0 and $\frac{1}{2}$ genotype A_1 will yield the three families $\{A_0\}$, $\{\tfrac{1}{2}A_0 + \tfrac{1}{2}A_1\}$, $\{A_1\}$ with probabilities $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$, respectively. The families $\{A_0\}$ and $\{A_1\}$ therefore each appear with frequency $\frac{1}{4} + \frac{1}{8}$ in the F_3 population, and the family $\{\tfrac{1}{2}A_0 + \tfrac{1}{2}A_1\}$ appears with frequency $\frac{1}{4}$. The total frequency of the gene A_j in the F_3 population is, of course, $\frac{1}{2} = \frac{1}{4} + \frac{1}{8} + \frac{1}{4} \cdot \frac{1}{2}$ so the F_3 mean = F_2 mean = F_1 mean. The total genetic variance V in the F_3 population may be partitioned into the three components

V(2,1;3) = variance among the means of the three F_3 family groups which trace back to three F_1 crosses $A_0 \times A_0$, $A_0 \times A_1$, and $A_1 \times A_1$, respectively,

V(3,2) = average variance among F_3 family means within F_2 families, and

V(3,3) = average variance among individual genotypes within F_3 families within F_2 families. Thus,

$$V(2,1;3) = \tfrac{1}{4}(a_0+a)^2 + \tfrac{1}{2}[\tfrac{1}{4}(a_0+a) + \tfrac{1}{2}\left\{\tfrac{1}{2}(a_0+a) + \tfrac{1}{2}(a_0-a)\right\} + \tfrac{1}{4}(a_0-a)]^2 + \tfrac{1}{4}(a_0-a)^2 - a_0^2$$

$$= \tfrac{1}{4}(a_0+a)^2 + \tfrac{1}{2}a_0^2 + \tfrac{1}{4}(a_0-a)^2 - a_0^2$$

$$= \tfrac{1}{2}a^2$$

$$V(3,2) = \tfrac{1}{4}(0) + \tfrac{1}{2}[\tfrac{1}{4}(a_0+a)^2 + \tfrac{1}{2}\left\{\tfrac{1}{2}(a_0+a) + \tfrac{1}{2}(a_0-a)\right\}^2 + \tfrac{1}{4}(a_0-a)^2 - a_0^2] + \tfrac{1}{4}(0)$$

$$= \tfrac{1}{2} \cdot \tfrac{1}{2}a^2$$

$$V(3,3) = \tfrac{3}{8}(0) + \tfrac{1}{4}\left\{\tfrac{1}{2}(a_0+a)^2 + \tfrac{1}{2}(a_0-a)^2 - a_0^2\right\} + \tfrac{3}{8}(0)$$

$$= \tfrac{1}{4}a^2.$$

Thus, $V = V(2,1) + V(3,2) + V(3,3) = \tfrac{1}{2}a^2 + \tfrac{1}{4}a^2 + \tfrac{1}{4}a^2 = a^2.$

The relationship between genotypic values of the F_3 families and their F_2 parents may be evaluated as the covariance of the F_3 family group means and the F_2 family means. Denoting this covariance by $Cov(2,1;3,2)$ we have

$$Cov(2,1;3,2) = \tfrac{1}{4}(a_0+a)(a_0+a)$$

$$+ \tfrac{1}{2}[\tfrac{1}{2}(a_0+a) + \tfrac{1}{2}(a_0-a)][\tfrac{1}{4}(a_0+a) + \tfrac{1}{2}\left\{\tfrac{1}{2}(a_0+a) + \tfrac{1}{2}(a_0-a)\right\} + \tfrac{1}{4}(a_0-a)]$$

$$+ \tfrac{1}{4}(a_0-a)(a_0-a)$$

$$- a_0a_0$$

$$= \tfrac{1}{4}(a_0+a)^2 + \tfrac{1}{2}(a_0^2) + \tfrac{1}{4}(a_0-a)^2 - a_0^2$$

$$= \tfrac{1}{2}a^2.$$

To partition the variance in a general generation r we note that this generation contains a collection of individuals all of which trace back to the same genotypic cross in generation k-1, $k \leq r$. We call such a collection of

individuals a (k,r)-family. Under the assumption of no linkage the genotypic distribution within a (k,r)-family will be unaltered by the random mating of full sibs; i.e., the genotypic distribution among individuals in generation r which trace back to a particular cross in generation (k-1) and the genotypic distribution among individuals in generation s which trace back to this same cross in generation k-1 will be identical. The variance V(k,k-1; r) among (k,r)-family means within (k-1,r)-families will therefore be independent of r and hence equal to V(k,k-1), the variance among (k,k)-family means within (k-1,k)-families. Furthermore, because corresponding (k,r)- and (k,s)-family means are identical then the covariance Cov(k,k-1;r,s) of (k,r)- and (k,s)-family means within (k-1) lineages will be equal to the variance V(k,k-1). Thus, we see in the above computations $V(2,1;3) = V(2,1) = Cov(2,1;3,2)$.

To compute the variance V(k,k-1) for a general k we note that in generation F_{k-1} the three genetically different families $\left\{A_0\right\}, \left\{\frac{1}{2}A_0 + \frac{1}{2}A_1\right\}, \left\{A_0\right\}$ appear with frequencies $\frac{1}{2}(1 - \frac{1}{2^{k-2}})$, $\frac{1}{2^{k-2}}$, $\frac{1}{2}(1 - \frac{1}{2^{k-2}})$, respectively. In generation F_k, then, the average genetic variance among family group means within F_{k-1} families is

$$V(k,k-1) = \tfrac{1}{2}(1 - \tfrac{1}{2^{k-2}})(0) + \tfrac{1}{2^{k-2}}\left[\tfrac{1}{4}(a_0+a)^2 + \tfrac{1}{2}\left\{\tfrac{1}{2}(a_0+a) + \tfrac{1}{2}(a_0-a)\right\}^2 + \tfrac{1}{4}(a_0-a)^2 - a_0^2\right]$$

$$+ \tfrac{1}{2}(1 - \tfrac{1}{2^{k-2}})(0)$$

$$= \frac{1}{2^{k-2}} \cdot \tfrac{1}{2}a^2$$

$$= \frac{1}{2^{k-1}}\, a^2$$

Since the total variance $V = a^2$ in generation F_k is

$$V = \sum_{j=2}^{k} V(j,j-1) + V(k,k)$$

we have that the average genetic variance within families in generation F_k is

$$V(k,k) = V - \sum_{j=2}^{k} V(j,j-1)$$

$$= a^2\left[1 - \frac{1}{2}\sum_{j=2}^{k}\frac{1}{2^{j-2}}\right]$$

$$= a^2\left[1 - \left(1 - \frac{1}{2^{k-1}}\right)\right]$$

$$= \frac{1}{2^{k-1}}a^2$$

n Unlinked Loci

If the original parents differ genetically at n loci then the F_1 population will consist of 2^n genotypes, each of which appears with frequency $\frac{1}{2^n}$. Denoting the two alleles at the i'th locus by A_0^i and A_1^i we write the general genotype as

$$\prod_{i=1}^{n} A_{j_i}^i$$

and use the same symbol to denote genotypic value. The identity (1), with the symbols denoting genotype rather than genotypic value becomes, for n loci,

$$\prod_{i=1}^{n} A_{j_i}^i = \prod_{i=1}^{n}\left[a_{0i} + (1-2j_i)a_i\right]$$

$$= \sum_{\upsilon=0}^{n} \sum_{I_\upsilon \text{ in } (1,\ldots,n)} \prod_{i \text{ in } I_\upsilon} (1-2j_i) \prod_{i \text{ in } I_\upsilon} a_i \prod_{i \text{ in } \bar{I}_\upsilon} a_{0i}$$

where $I_\upsilon = (i_1,\ldots,i_\upsilon)$ is a set of υ integers with $1 \le i_1 < \ldots < i_\upsilon \le n$ and $\bar{I}_\upsilon$ is the complement of I_υ with respect to the set $(1,\ldots,n)$; i.e., $\bar{I}_\upsilon$ is the set of integers in $(1,\ldots,n)$ which are not in I_υ. The notation

$$\sum_{I_\upsilon \text{ in } (1,\ldots,n)}$$

then means the sum over all of the $\binom{n}{\upsilon}$ subsets of $(1,\ldots,n)$ containing υ elements, and

$$\prod_{i \text{ in } I_\upsilon} x_i = x_{i_1} \cdot x_{i_2} \cdots x_{i_\upsilon}$$

while the notation

$$\prod_{i \text{ in } \bar{I}_\upsilon} y_i$$

means the product of the $n-\upsilon$ terms y_i which satisfy i in $\bar{I}_\upsilon$. Upon expanding the expression

$$\prod_{i \text{ in } I_\upsilon} a_i \prod_{i \text{ in } \bar{I}_\upsilon} a_{0i} = \prod_{i \text{ in } I_\upsilon} \frac{A_0^i - A_1^i}{2} \prod_{i \text{ in } \bar{I}_\upsilon} \frac{A_0^i + A_1^i}{2}$$

$$= \sum_{\alpha=0}^{\upsilon} \sum_{I_\alpha \text{ in } I_\upsilon} \sum_{\beta=0}^{n-\upsilon} \sum_{I_\beta \text{ in } \bar{I}_\upsilon} (-1)^{\upsilon-\alpha} \prod_{I_\alpha+I_\beta} A_0^i \prod_{I_\upsilon-I_\alpha+\bar{I}_\upsilon-I_\beta} A_1^i$$

and replacing the genotype

$$\prod_{I_\alpha+I_\beta} A_0^i \qquad \prod_{I_\upsilon-I_\alpha+\bar{I}_\upsilon-I_\beta} A_1^i$$

by its genotypic value we then define the <u>genotypic effect</u> g_{I_υ} by

$$g_{I_\upsilon} = \sum_{\alpha=0}^{\upsilon} \sum_{I_\alpha \text{ in } I_\upsilon} \sum_{\beta=0}^{n-\upsilon} \sum_{I_\beta \text{ in } \bar{I}_\upsilon} (-1)^{\upsilon-\alpha} \prod_{I_\alpha+I_\beta} A_0^i \prod_{I_\upsilon-I_\alpha+\bar{I}_\upsilon-I_\beta} A_1^i$$

Then we may write for the general genotypic value $\prod_{i=1}^{n} A_{j_i}^i$

$$\prod_{i=1}^{n} A_{j_i}^i = \prod_{I_n} A_{j_i}^i = \sum_{\upsilon=0}^{n} \sum_{I_\upsilon \text{ in } I_n} \prod_{I_\upsilon} (1-2j_i)\, g_{I_\upsilon}.$$

For example, for n = 2, and letting $A_0^1 = A$, $A_1^1 = a$, $A_0^2 = B$, $A_1^2 = b$, we have

$$g_0 = \frac{AB + Ab + aB + ab}{4}$$

$$g_1 = \frac{AB + Ab - aB - ab}{4}$$

$$g_2 = \frac{AB - Ab + aB - ab}{4}$$

$$g_{12} = \frac{AB - Ab - aB + ab}{4}$$

and we interpret the quantity g_1, for example, as the average effect of the gene A while $-g_1$ is the average effect of the gene a, and g_{12} is the epistatic effect. Thus, the table of genotypic values obtained from the formula

$$A^1_{j_1} A^2_{j_2} = g_0 + (1-2_{j_1})g_1 + (1-2_{j_2})g_2 + (1-2_{j_1})(1-2_{j_2})g_{12}$$

would read as follows:

	A	a
B	$g_0 + g_1 + g_2 + g_{12}$	$g_0 - g_1 + g_2 - g_{12}$
b	$g_0 + g_1 - g_2 - g_{12}$	$g_0 - g_1 - g_2 + g_{12}$

Now if we assume no linkage; i.e., independent segregation at each of the n loci, then the array of families in generation F_{k-1} under random mating among full sibs may be written

$$\prod_{I_n} \left[\frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right)\left\{A_0^i\right\} + \frac{1}{2^{k-2}}\left\{\frac{1}{2}A_0^i + \frac{1}{2}A_1^i\right\} + \frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right)\left\{A_1^i\right\}\right]$$

$$(3) \quad = \sum_{\upsilon=0}^{n} \sum_{I_\upsilon \text{ in } I_n} \sum_{\alpha=0}^{\upsilon} \sum_{I_\alpha \text{ in } I_\upsilon} \left[\frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right)\right]^{n-\upsilon+\alpha} \left[\frac{1}{2^{k-2}}\right]^{\upsilon-\alpha}$$

$$\cdot \left\{\frac{1}{2^{\upsilon-\alpha}} \sum_{\beta=0}^{\upsilon-\alpha} \sum_{I_\beta \text{ in } I_\upsilon - I_\alpha} \prod_{I_\alpha + I_\beta} A_0^i \prod_{(I_n - I_\upsilon) + (I_\upsilon - I_\alpha - I_\beta)} A_1^i\right\}$$

The F_{k-1} family defined by the brackets $\{\ \}$ in (3) contains $2^{\upsilon-\alpha}$ genotypes, each appearing with the frequency $\frac{1}{2^{\upsilon-\alpha}}$. Under random mating among full sibs this F_{k-1} family produces $3^{\upsilon-\alpha}$ families in generation F_k. Genetic variation, both within and between these $3^{\upsilon-\alpha}$ families occurs only at the $\upsilon-\alpha$ loci defined by the set $I_\upsilon - I_\alpha$, and this genetic variation is precisely of the form which would be obtained in an F_2 population derived from random mating in an F_1 population which exhibits genotypic variation at $\upsilon-\alpha$ independent loci. We may compute the variance among families in such an F_2 population from the formula (2) obtained earlier for the case n = 1. Thus

$$V_{I_\upsilon - I_\alpha}(2,1) = \prod_{I_\upsilon - I_\alpha}\left[\frac{1}{2}\left(\frac{A_i^0 - A_i^1}{2}\right)^2 + \left(\frac{A_i^0 + A_i^1}{2}\right)^2\right] - \prod_{I_\upsilon - I_\alpha}\left(\frac{A_i^0 + A_i^1}{2}\right)^2$$

$$= \sum_{\beta=1}^{\upsilon-\alpha} \sum_{I_\beta \text{ in } I_\upsilon - I_\alpha} \frac{1}{2^\beta}\left[\prod_{I_\beta}\left(\frac{A_i^0 - A_i^1}{2}\right) \prod_{I_\upsilon - I_\alpha - I_\beta}\left(\frac{A_i^0 + A_i^1}{2}\right)\right]^2 .$$

Hence, we may write V(k,k-1) as

$$V(k,k-1) = \sum_{\upsilon=1}^{n} \sum_{I_\upsilon} \sum_{\alpha=0}^{\upsilon} \sum_{I_\alpha \text{ in } I_\upsilon} \left[\frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right)\right]^{n-\upsilon+\alpha}\left[\frac{1}{2^{k-2}}\right]^{\upsilon-\alpha}$$

$$\cdot \sum_{\beta=1}^{\upsilon-\alpha} \sum_{I_\beta \text{ in } I_\upsilon - I_\alpha} \frac{1}{2^\beta}\left[\prod_{I_\alpha} A_0^i \prod_{I_n - I_\upsilon} A_1^i \prod_{I_\beta}\left(\frac{A_0^i - A_1^i}{2}\right) \prod_{I_\upsilon - I_\alpha - I_\beta}\left(\frac{A_0^i + A_1^i}{2}\right)\right]^2$$

or, letting $A_0^i = a_{i0} + a_i$, $A_1^i = a_{i0} - a_i$, so that

$$\prod_{I_\alpha} A_0^i \prod_{I_n - I_\upsilon} A_1^i \prod_{I_\beta}\left(\frac{A_0^i - A_1^i}{2}\right) \prod_{I_\upsilon - I_\alpha - I_\beta}\left(\frac{A_0^i + A_1^i}{2}\right)$$

$$= \prod_{I_\alpha}(a_{i0}+a_i) \prod_{I_n - I_\upsilon}(a_{i0}-a_i) \prod_{I_\beta} a_i \prod_{I_\upsilon - I_\alpha - I_\beta} a_{i0}$$

$$= \sum_{\gamma=0}^{\alpha} \sum_{I_\gamma \text{ in } I_\alpha} \sum_{\delta=0}^{n-\upsilon} \sum_{I_\delta \text{ in } I_n - I_\upsilon} (-1)^{\delta} \prod_{I_\beta + I_\gamma + I_\delta} a_i \prod_{(I_\alpha - I_\gamma) + (I_n - I_\upsilon - I_\delta) + (I_\upsilon - I_\alpha - I_\beta)} a_{i0}$$

$$= \sum_{\gamma=0}^{\alpha} \sum_{I_\gamma \text{ in } I_\alpha} \sum_{\delta=0}^{n-\upsilon} \sum_{I_\delta \text{ in } I_n - I_\upsilon} (-1)^{\delta} g_{I_\beta + I_\gamma + I_\delta}$$

then

$$V(k,k-1) = \sum_{\upsilon=1}^{n} \sum_{I_\upsilon} \sum_{\alpha=0}^{\upsilon} \sum_{I_\alpha \text{ in } I_\upsilon} \left[\frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right)\right]^{n-\upsilon+\alpha} \left[\frac{1}{2^{k-2}}\right]^{\upsilon-\alpha}$$

(4)
$$\sum_{\beta=1}^{\upsilon-\alpha} \sum_{I_\beta \text{ in } I_\upsilon - I_\alpha} \frac{1}{2^{\beta}} \left[\sum_{\gamma=0}^{\alpha} \sum_{I_\gamma \text{ in } I_\alpha} \sum_{\delta=0}^{n-\upsilon} \sum_{I_\delta \text{ in } I_n - I_\upsilon} (-1)^{\delta} g_{I_\beta + I_\gamma + I_\delta} \right]^2$$

We shall show that this expression (4) for genetic variance among F_k family means within F_{k-1} families reduces to

(5)
$$V(k,k-1) = \sum_{m=1}^{n} \left[\left(1 - \frac{1}{2^{k-1}}\right)^m - \left(1 - \frac{1}{2^{k-2}}\right)^m \right] \sum_{I_m} g^2_{I_m}$$

To accomplish this we compute the coefficient of the cross product

$$g_{I_\beta + I_\gamma + I_\delta} \; g_{I_\beta + I_{\gamma'} + I_{\delta'}}$$

in (4). Since

$$I_\beta \text{ in } I_\upsilon - I_\alpha$$

$$I_\gamma \text{ in } I_\alpha \text{ in } I_\upsilon$$

$$I_\delta \text{ in } I_n - I_\upsilon$$

then

$$(I_\beta + I_\gamma + I_\delta) \cap (I_\beta + I_{\gamma'} + I_{\delta'}) = I_\beta + I_\gamma \cap I_{\gamma'} + I_\delta \cap I_{\delta'}$$

$$= (i_1, \ldots, i_m), \; 1 \le i_1 < \ldots < i_m \le n$$

$$= I_m, \text{ say,}$$

where $I_x \cap I_y$ is the set of elements common to the two sets I_x and I_y. Let

$$I_q = I_\beta + I_\gamma + I_\delta - I_m$$

$$I_{q'} = I_\beta + I_{\gamma'} + I_{\delta'} - I_m$$

Now consider the particular case of $g_{I_\beta + I_\gamma + I_\delta}\, g_{I_\beta + I_{\gamma'} + I_{\delta'}}$ where

$$I_\beta = (j_1,\ldots,j_{m-x}),\ I_\beta \text{ in } I_m,\ i_1 \leq j_1 < \ldots < j_{m-x} \leq i_m$$

$$I_\gamma \cap I_{\gamma'} = (j_{m-x+1},\ldots,j_{m-x+r}),\ i_1 \leq j_{m-x+1} < \ldots < j_{m-x+r} \leq i_m$$

$$I_\delta \cap I_{\delta'} = (j_{m-x+r+1},\ldots,j_m),\ i_1 \leq j_{m-x+r+1} < \ldots < j_m \leq i_m$$

and let I_{γ_0}, I_{δ_0}, $I_{\gamma'_0}$, $I_{\delta'_0}$ be sets which satisfy

(6) $$(j_{m-x+1},\ldots,j_{m-x+r}) \text{ in } I_{\gamma_0} \text{ in } I_q + (j_{m-x+1},\ldots,j_{m-x+r})$$

$$I_{\delta_0} = I_\gamma + I_\delta - I_{\gamma_0}$$

$$(j_{m-x+1},\ldots,j_{m-x+r}) \text{ in } I_{\gamma'_0} \text{ in } I_{q'} + (j_{m-x+1},\ldots,j_{m-x+r})$$

$$I_{\delta'_0} = I_{\gamma'} + I_{\delta'} - I_{\gamma'_0}$$

To compute the coefficient of $g_{I_\beta + I_{\gamma_0} + I_{\delta_0}}\, g_{I_\beta + I_{\gamma'_0} + I_{\delta'_0}}$ in (4) we note that since I_γ in I_α in I_υ and I_β in $I_\upsilon - I_\alpha$ in I_υ then I_υ must satisfy

(7) $$(j_1,\ldots,j_{m-x+r}) + \left\{ I_{\gamma_0} - (j_{m-x+1},\ldots,j_{m-x+r}) \right\} + \left\{ I_{\gamma'_0} - (j_{m-x+1},\ldots,j_{m-x+r}) \right\} \text{ in } I_\upsilon$$

and since I_δ in $I_n - I_\upsilon$ then I_υ must not contain I_{δ_0} or $I_{\delta'_0}$; i.e.,

$$I_q + I_{q'} + (j_{m-x+r+1},\ldots,j_m) - \left\{ I_{\gamma_0} - (j_{m-x+1},\ldots,j_{m-x+r}) \right\}$$

(8)
$$- \left\{ I_{\gamma_0'} - (j_{m-x+1},\ldots,j_{m-x+r}) \right\} \text{ in } I_n - I_\upsilon$$

Likewise, since I_γ in I_α then I_α must satisfy

$$(j_{m-x+1},\ldots,j_{m-x+r}) + \left\{ I_{\gamma_0} - (j_{m-x+1},\ldots,j_{m-x+r}) \right\}$$

(9)
$$+ \left\{ I_{\gamma_0'} - (j_{m-x+1},\ldots,j_{m-x+r}) \right\} \text{ in } I_\alpha$$

and since I_β in $I_\upsilon - I_\alpha$ then I_α must satisfy

(10) $$(j_1,\ldots,j_{m-x}) \text{ in } I_\upsilon - I_\alpha$$

We therefore have, as the coefficient of $g_{I_\beta + I_{\gamma_0} + I_{\delta_0}}\, g_{I_\beta + I_{\gamma_0'} + I_{\delta_0'}}$ in (4)

$$\sum_{\upsilon = m-x-r+\gamma_0+\gamma_0'}^{n-q-q'-x-r+\gamma_0+\gamma_0'} \quad \sum_{\substack{I_\upsilon \text{ satisfying} \\ (7) \text{ and } (8)}} \quad \sum_{\alpha = \gamma_0+\gamma_0'-r}^{\upsilon-(m-x)} \quad \sum_{\substack{I_\alpha \text{ satisfying} \\ (9) \text{ and } (10)}}$$

$$\cdot \left[\frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right) \right]^{n-\upsilon+\alpha} \left[\frac{1}{2^{k-2}} \right]^{\upsilon-\alpha} \frac{1}{2^{m-x}} (-1)^{\delta_0+\delta_0'}$$

multiplied by 2 if $I_\beta + I_{\gamma_0} + I_{\delta_0} \neq I_\beta + I_{\gamma_0'} + I_{\delta_0'}$. Letting

$$C_k = \frac{\frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right)}{\frac{1}{2^{k-2}}}$$

this coefficient becomes

$$(11)\quad \sum_{\upsilon = m-x-r+\gamma_0+\gamma_0'}^{n-q-q'-x-r+\gamma_0+\gamma_0'} \binom{n-m-q-q'}{\upsilon-m-x+r-\gamma_0-\gamma_0'} \sum_{\alpha = \gamma_0+\gamma_0'-r}^{\upsilon-(m-x)} \binom{\upsilon-m+x+r-\gamma_0-\gamma_0'}{\alpha-\gamma_0-\gamma_0'+r}$$

$$\cdot C_k^{\alpha} C_k^{-\upsilon} \left[\frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right)\right]^n \frac{1}{2^{m-x}}(-1)^{\delta_0+\delta_0'}$$

$$= \sum_{\upsilon = m-x-r+\gamma_0+\gamma_0'}^{n-q-q'-x-r+\gamma_0+\gamma_0'} \binom{n-m-q-q'}{\upsilon-m+x+r-\gamma_0-\gamma_0'} \left(\frac{1+C_k}{C_k}\right)^{\upsilon-m+x+r-\gamma_0-\gamma_0'}$$

$$\cdot C_k^{-(m-x)} \frac{1}{2^{m-x}}\left[\frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right)\right]^n (-1)^{\delta_0+\delta_0'}$$

$$= (1 + 2C_k)^{n-m-q-q'} C_k^{-n+m+q+q'-(m-x)} \frac{1}{2^{m-x}}\left[\frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right)\right]^n (-1)^{\delta_0+\delta_0'}$$

$$= \left[\frac{1}{2}\left(1 - \frac{1}{2^{k-2}}\right)\right]^{q+q'+x} \left[\frac{1}{2^{k-2}}\right]^{m-x} \frac{1}{2^{m-x}} (-1)^{\delta_0+\delta_0'}$$

which is seen to be independent of r. We now observe that for the fixed sets

$$I_\beta = (j_1, \ldots, j_{m-x})$$

$$I_\gamma \cap I_{\gamma'} = (j_{m-x+1}, \ldots, j_{m-x+r})$$

$$I_\delta \cap I_{\delta'} = (j_{m-x+r+1}, \ldots, j_m)$$

the number of different sets $I_{\gamma_0^*}$ which satisfy (6) and contain exactly γ_0 elements is $\binom{\gamma_0}{r}\binom{q}{r}$. Since $I_\beta + I_{\gamma_0^*} + I_{\delta_0^*} = I_\beta + I_{\gamma_0} + I_{\delta_0}$ then for all such $I_{\gamma_0^*}$, $I_{\delta_0^*} = I_\gamma + I_\delta - I_{\gamma_0^*}$, we have

$$(12)\quad g_{I_\beta + I_{\gamma_0^*} + I_{\delta_0^*}} g_{I_\beta + I_{\gamma_0'} + I_{\delta_0'}} = g_{I_\beta + I_{\gamma_0} + I_{\delta_0}} g_{I_\beta + I_{\gamma_0'} + I_{\delta_0'}}$$

and the coefficient (11). Furthermore, since (11) depends only upon the number $\delta_0 = q + x - \gamma_0$ of elements in the set I_{δ_0} and not upon the set itself then permitting δ_0 to range over the interval $x - r \leq \delta_0 \leq q + x - r$ we obtain as the coefficient

$$\left[\frac{1}{2}(1 - \frac{1}{2^{k-2}})\right]^{q+q'+x} \left[\frac{1}{2^{k-1}}\right]^{m-x} (-1)^{\delta_0+x-r} \sum_{\gamma_0=r}^{q+r} \binom{q}{\gamma_0-r} (-1)^{q-\gamma_0+r}$$

which vanishes when $q>0$. This proves, then, that when $I_\beta + I_\gamma + I_\delta \neq I_\beta + I_{\gamma'} + I_{\delta'}$ the coefficient of $g_{I_\beta + I_\gamma + I_\delta}\, g_{I_\beta + I_{\gamma'} + I_{\delta'}}$ in (4) is 0.

We have now to examine the remaining case $q = q' = 0$; i.e., we shall now compute the coefficient of $g^2_{I_\beta + I_\gamma + I_\delta}$ in (4). For this case the coefficient (11) becomes

$$\left[\frac{1}{2}(1 - \frac{1}{2^{k-2}})\right]^{x} \left[\frac{1}{2^{k-1}}\right]^{m-x} \tag{13}$$

Since this does not depend upon r and since, for each r, $0 \leq r \leq x$, there are $\binom{x}{r}$ sets $I_\gamma \cap I_{\gamma'}$ which can be formed by choosing r elements from the set $(j_{m-x+1},\ldots,j_m)$ then permitting r to vary over the interval $0 \leq r \leq x$ we obtain

$$\sum_{r=0}^{x} \binom{x}{r} = 2^x$$

repetitions of (13). Furthermore, there are $\binom{m}{x}$ sets $(j_{m-x+1},\ldots,j_m)$ which can be formed by choosing x elements from the set $(i_1,\ldots,i_m)$, so we obtain as the coefficient of $g^2_{I_\beta + I_\gamma + I_\delta = I_m}$, $1 \leq m \leq n$,

$$\sum_{x=0}^{m-1} \binom{m}{x} 2^x \left[\frac{1}{2}(1 - \frac{1}{2^{k-2}})\right]^{x} \left[\frac{1}{2^{k-1}}\right]^{m-x}$$

$$= (1 - \frac{1}{2^{k-2}} + \frac{1}{2^{k-1}})^m - (1 - \frac{1}{2^{k-2}})^m$$

$$= (1 - \frac{1}{2^{k-1}})^m - (1 - \frac{1}{2^{k-2}})^m$$

and (5) is established.

The total genetic variance in the F_k population is given by *

$$V = \sum_{m=1}^{n} \sum_{I_m} g^2_{I_m}$$

$$= V(k,k) + \sum_{j=2}^{k} V(j,j-1)$$

so that

$$V(k,k) = V - \sum_{j=2}^{k} V(j,j-1)$$

where

$$\sum_{j=2}^{k} V(j,j-1) = \sum_{m=1}^{n} \left[\sum_{x=0}^{m} (-1)^x \binom{m}{x} \sum_{j=2}^{k} \left\{ \frac{1}{2^{x(j-1)}} - \frac{1}{2^{x(j-2)}} \right\} \right] \sum_{I_m} g^2_{I_m}$$

$$= \sum_{m=1}^{n} \left[\sum_{x=0}^{m} (-1)^x \binom{m}{x} \left\{ \frac{1}{2^{x(k-1)}} - 1 \right\} \right] \sum_{I_m} g^2_{I_m}$$

$$= \sum_{m=1}^{n} \left(1 - \frac{1}{2^{k-1}}\right)^m \sum_{I_m} g^2_{I_m}$$

Thus,

$$V(k,k) = \sum_{m=1}^{n} \left[1 - \left(1 - \frac{1}{2^{k-1}}\right)^m \right] \sum_{I_m} g^2_{I_m} .$$

Discussion and Summary

We consider the problem of estimating genetic variance components in a haploid population derived from the cross of two genotypes which differ at n loci. The notation

$$\prod_{i=1}^{n} A^i_{j_i}, \quad j_i = 0 \text{ or } 1$$

is used to denote either genotype or genotypic value. The epistatic, or interaction effect of the m loci $i_1, i_2, \ldots, i_m$ in the set $I_m = (i_1, i_2, \ldots, i_m)$, $1 \leq i_1 < \ldots < i_m \leq n$, is defined as

*See [2].

$$g_{I_m} = \sum_{\alpha=0}^{m} \sum_{I_\alpha \text{ in } I_m} \sum_{\beta=0}^{n-m} \sum_{I_\beta \text{ in } I_m - I_\alpha} (-1)^{m-\alpha} \prod_{I_\alpha + I_\beta} A_0^i \prod_{I_n - I_\alpha - I_\beta} A_1^i$$

We assume that segregation is independent at n loci and that full sibs are mated at random in the successive generations following the initial cross. The genotypic distribution is therefore constant through successive generations and has a mean value of $g_{I_0} = \mu$ and variance

$$V = \sum_{m=1}^{n} \sum_{I_m} g_{I_m}^2$$

or, denoting

$$\sigma_m^2 = \sum_{I_m} g_{I_m}^2 ,$$

then

$$V = \sum_{m=1}^{n} \sigma_m^2 .$$

Successive generations exhibit a family structure and in generation r there exists a collection of individuals, called a (k,r)-family, which all trace back to the same genotypic cross in generation k-1, $k \leq r$. Under the assumption of no linkage the (k,r)-family means and the (k,s)-family means are identical; their covariance, written Cov(k;r,s), is therefore equal to the variance V(k) among (k,k)-family means. The variance V(k) is expressible as a linear combination of the genetic variance components σ_m^2, m = 1,...,n; thus,

$$\text{Cov}(k;r,s) = V(k) = \sum_{m=1}^{n} \left(1 - \frac{1}{2^{k-1}}\right)^m \sigma_m^2 .$$

Furthermore, the variance V(k) and, likewise, the covariance Cov(k; r,s) may be partitioned into k-1 components V(j,j-1), j=2,...,k, representing the average variance among (j,k)-family means within (j-1,k)-families, where

$$\text{Cov}(j,j-1;r,s) = V(j,j-1) = \sum_{m=1}^{n} \left[\left(1 - \frac{1}{2^{k-1}}\right)^m - \left(1 - \frac{1}{2^{k-2}}\right)^m \right] \sigma_m^2 .$$

The average genetic variance within (k,k)-families is then

$$\begin{aligned} V(k,k) &= V - V(k) \\ &= \sum_{m=1}^{n} \left[1 - \left(1 - \frac{1}{2^{k-1}}\right)^m \right] \sigma_m^2 . \end{aligned}$$

It should be pointed out that the variance and covariance formulae given here apply to the population and not directly to an experimental situation. We may illustrate the application of these results by considering an experiment in which the k'th experimental generation is composed of t_1 individuals per (k,k)-experimental family and t_{k-j+2}(j,k)-experimental families per (j-1,k)-experimental family, j=2,...,k, arranged in a completely randomized design. Then (see [1]) the expectation of the experimental variance $\hat{V}(j,j-1)$ is given by

$$(\prod_{i=1}^{k-1} t_i)E\left\{\hat{V}(j,j-1)\right\} = \bar{\sigma}_e^2 + V(k,k) + \sum_{j=2}^{k} V(j,j-1) \prod_{i=1}^{k-j+1} t_i$$

where $\bar{\sigma}_e^2$ is the average environmental variance.

References

[1] Robson, D. S. Applications of the k_4-statistic to genetic variance component analysis. Unpublished mimeograph. 1955.

[2] Robson, D. S. A factorial analysis of variance model and its application in genetics. BU-64-M, April, 1956.